Moment-based estimation of nonlinear regression models under unobserved heterogeneity, with applications to non-negative and fractional responses*

Esmeralda A. Ramalho

Department of Economics and CEFAGE-UE, Universidade de Evora

Joaquim J.S. Ramalho

Department of Economics and CEFAGE-UE, Universidade de Evora

July 2012

Abstract

In this paper we suggest simple moment-based estimators to deal with unobserved heterogeneity in nonlinear regression models that treat observed and omitted covariates in a similar manner. The results derived in the paper apply to a class of regression models that includes as particular cases exponential and logit and complementary loglog fractional regression models. Unlike previous approaches, which typically require distributional assumptions on the unobservables, a conditional mean assumption is enough for consistent estimation of the structural parameters. Under the additional assumption that the dependence between observables and unobservables is restricted to the conditional mean, consistent estimation of partial effects conditional only on the former is also possible without making distributional assumptions on the latter.

Keywords: unobserved heterogeneity, endogeneity, fractional regression, exponential regression, transformation regression models.

JEL Classification: C25, C51, C26.

*Financial support from Fundacao para a Ciencia e a Tecnologia (grant PTDC/EGE-ECO/119148/2010) is gratefully acknowledged. We are also indebted for helpful comments to Joao Santos Silva. Address for correspondence: Joaquim J.S. Ramalho, Department of Economics, Universidade de Évora, Largo dos Colegiais, 7000-803 ÉVORA, Portugal (e-mail: jsr@uevora.pt).

1 Introduction

Economic theory often postulates that a response variable depends on both observed and unobserved individual factors; see *inter alia* Heckman (2000, 2001), and his notion of a Marshallian structural function, and Wooldridge (2005), and his related concept of a structural expectation of interest. Therefore, empirical researchers often have to deal with the so-called problem of 'omitted variables' or 'unobservables' in their econometric models. When the model is linear in the parameters, i.e. both observables and unobservables enter the model additively, this issue is easily dealt with, irrespective of how excluded and included covariates are related. For example, if the omitted variables are uncorrelated with the variables included in the model, then unobservables may be simply ignored and standard application of ordinary least squares (OLS) produces unbiased estimators; if, instead, unobserved and observed covariates are correlated, then, provided that a set of instruments is available for the endogenous regressors, instrumental variables based approaches, such as the generalized method of moments (GMM), may be applied to obtain consistent estimators of the parameters of interest.

While linear models are widely used in econometrics, there are many circumstances where it is preferable to specify a nonlinear regression model. A prominent example of this practice occurs when the variable of interest, y, has a bounded nature. In such a case, linear specifications provide, in general, an inadequate description of the behavior of y, because they do not impose any restriction on the range of values yielded by the structural function or expectation relating y to the observables x and the unobservables u. For example, let $E(y|x,u)$ be the conditional expectation of y given x and u. If all realizations of y are nonnegative or fractional, then $E(y|x,u) > 0$ or $0 < E(y|x,u) < 1$, respectively, while the linear model implicitly assumes $-\infty < E(y|x,u) < \infty$. Unfortunately, in the framework of nonlinear regression models it is much more complicated to deal with unobserved heterogeneity. Indeed, if not properly accounted for, the omission of explanatory variables in nonlinear regression models will generally lead to inconsistent estimation of the regression coefficients and other quantities of interest, even in the case of neglected heterogeneity.[1] For example, Ramalho and Ramalho (2010) found the following consequences of neglected heterogeneity in the context of binary and fractional regression models: (i) it produces an attenuation bias in the estimation of regression coefficients; (ii) apart from some particular cases, it generates biased estimation of (averaged across the distribution of unobservables) partial effects; and (iii) although innocuous for the size of Wald tests for the significance of observed regressors, it substantially reduces their power.

[1] We use the term 'neglected heterogeneity' to designate the case where the unobserved and the included covariates are independent.

Despite the general acknowledgement of the deleterious effects of unobserved heterogeneity in nonlinear specifications, many empirical researchers still specify nonlinear models that simply do not allow for unobservables as if ignoring the problem would eliminate it. Other common practice in empirical work (which, in practical terms, is identical to the previous approach) is the introduction of heterogeneity in the model in such a way that it can immediately be discarded again, i.e., observables and unobservables are treated, often without any plausible reason, in a non-symmetrical way, just to ensure the separability of the observable and unobservable components. Authors that do incorporate the heterogeneity in the model in a sensible manner then typically choose one of the following strategies: (i) make strong distributional assumptions for the unobservables, which often generate poorly fitting models; or (ii) work with linearized versions of the model of interest, which, typically, cannot (or should not) be applied in cases where boundary values of y are observed with nonzero probability (e.g. the value zero of nonnegative outcomes in log-transformed models).

In this paper we propose another solution to deal with neglected heterogeneity and/or endogeneity issues in the framework of a particular class of nonlinear regression models that treat observed and omitted covariates in a similar manner. Like strategy (ii), we suggest applying a transformation to the specified nonlinear structural model that makes dealing with unobservables a much simpler task in the class of models considered. Unlike strategy (ii), the proposed transformation regression model accommodates values of y observed at one or both of its boundaries in many cases (e.g. the value zero of nonnegative outcomes; the value zero or one of fractional response variables). The suggested model generates a set of orthogonality conditions between the (transformed) error term and the explanatory (or instrumental) variables and may be easily estimated by GMM. Therefore, its implementation is typically much simpler than that of strategy (i), where often the parameters have to be estimated using simulation techniques. Moreover, unlike strategy (i), no distributional assumptions are required, a conditional mean assumption regarding the unobservables being enough for consistent estimation of the parameters of interest. In the case of endogenous regressors, another advantage of our method is that no reduced form for those covariates needs to be specified, although such information may be also incorporated in the estimation process.

The focus of the paper is on the estimation of the parameters that appear in the structural model, which are of interest in its own right for general policy analysis or for testing restrictions imposed by economic theory, for example. However, when dealing with unobserved heterogeneity, empirical researchers may be also interested in computing partial effects conditional only on observables. Actually, some authors, notably Wooldridge (2005), argue that in the presence of

unobservables, the quantities of primary interest for empirical analysis are often not so much the structural parameters but rather the partial effects averaged across the population distribution of any unobserved heterogeneity. Interestingly, as shown by Wooldridge (2005), under appropriate distributional assumptions on the unobservables, there are several examples of nonlinear regression models where it is possible to estimate consistently those partial effects, even though the structural parameters are not identified. Therefore, in this paper we consider also the estimation of partial effects conditional only on observables and show how, after obtaining consistent estimates for the structural parameters using the proposed transformation regression model, it is also possible to estimate consistently those quantities under the additional assumption that the dependence between observables and unobservables is restricted to the conditional mean.

Throughout the paper, all results are first derived for a general class of nonlinear, single-index regression models and then specialized for the exponential regression model (Wooldridge, 1992) and various alternative fractional regression models (Papke and Wooldridge, 1996). In the former case, several estimators already known in the econometric literature are produced. In the latter case, new estimators for dealing with neglected heterogeneity and endogeneity issues are suggested. To illustrate the advantages of our estimators over existing approaches we use Monte Carlo methods. An empirical application concerning corporate capital structure decisions, where the variable of interest has a fractional nature, is also provided.

This paper is organized as follows. Section 2 discusses briefly the specification of nonlinear structural models and provides a brief review of GMM estimation. Section 3 describes the proposed transformation regression model. Section 4 considers estimation of partial effects conditional on observables. Section 5 is dedicated to the Monte Carlo simulation study. Section 6 presents the empirical application. Finally, Section 7 concludes.

2 Specification and estimation of nonlinear regression models

2.1 Structural model

Let y be an observed (limited) dependent variable and let x and q be k and c vectors of observed explanatory variables and unobserved heterogeneity, respectively. Assume that x contains a constant term and denote by θ and η the vectors of parameters associated to x and q, respectively. Without loss of generality, let $u \equiv q\eta$.

Throughout this paper we assume that there is no specification error of any sort so that u represents basically the inherent randomness in human behavior, reflecting actual differences in

actions, tastes, technologies, etc. across the sampled economic agents, rather than measurement, sampling or other specification errors. We assume also that economic theory implies restrictions on the structure of the model, generating a nonlinear single-index representation for the relationship between y and (x, u). Thus, we consider what Heckman (2000) calls a 'well-posed economic model', that is 'a model that specifies all of the input processes, observed and unobserved by the analyst, and their relationship to outputs'. Under these assumptions, the resultant behavioral or structural model, called by Heckman (2000, 2001) a Marshallian causal function, may be specified as:

$$y = G(x\theta + u), \tag{1}$$

where $G(\cdot)$ is a known nonlinear function that imposes the bounded nature of y on the model.[2] We assume that $G(\cdot)$ is strictly monotonic, continuously differentiable and not additively separable.

From (1), it follows that

$$E(y|x) = E_u[G(x\theta + u)] = \int_{\mathcal{U}} G(x\theta + u) f(u|x)\,du, \tag{2}$$

where $E_u[\cdot]$ denotes expectation with respect to the conditional distribution of u and $\mathcal{U}$ and $f(u|x)$ denote, respectively, the sample space and the conditional (on the observables) density of u, which in this case, for simplicity, is assumed to be a scalar. Equation (2) shows that conditioning on the observed explanatory variables does not remove, in general, the dependency of the model on unobservables. However, see Example 1 below for a well known exception.

To illustrate the main points of the paper, all results will be specialized to the following two models.

Example 1 (Exponential regression model) *This model, which is commonly used to describe nonnegative outcomes, may be expressed as*

$$G(x\theta + u) = \exp(x\theta + u);$$

see, for example, Wooldridge (1992). As $G(\cdot)$ is multiplicatively separable in terms of x and u, it follows that $E(y|x) = \exp(x\theta)$ provided that $E_u[\exp(u)|x] = 1$, which is a common assumption in this framework. Therefore, conventional application of quasi-maximum likelihood (QML) methods yields consistent estimators for θ; see Santos Silva and Tenreyo (2006) for alternative QML estimators for exponential regression models.

[2] See Heckman (2000) for a rigorous definition of Marshallian causal functions.

Example 2 (Fractional regression models) *Models for variables defined on the unit interval ($0 \leq y \leq 1$) were first suggested by Papke and Wooldridge (1996). In this context, some popular choices for $G(\cdot)$ are the probit, logit and complementary loglog functional forms described in Table 1. In contrast to the previous example, the terms involving x and u are not directly separable in any of those models. Therefore, as discussed by Ramalho and Ramalho (2010), even under neglected heterogeneity does the Bernoulli-based QML method usually applied in this framework yield inconsistent estimators for θ.*

Table 1 about here

2.2 Features of interest

Estimating θ in (1) or (2) is typically a challenging task given the presence of unobservables in both expressions. Moreover, conditioning the analysis on both observables and unobservables, as in (1), or only on observables, as in (2), has very distinct implications for econometric studies. Indeed, in the former case the partial effects of unitary changes in a continuous covariate x_l on y are given by

$$\frac{\partial y}{\partial x_l} = \theta_l g(x\theta + u), \tag{3}$$

where $g(\cdot)$ is the derivative of $G(\cdot)$, while in the latter case the partial effects are given by

$$\frac{\partial E(y|x)}{\partial x_l} = \frac{\partial E_u[G(x\theta + u)]}{\partial x_l} = \int_{\mathcal{U}} \frac{\partial [G(x\theta + u)]}{\partial x_l} f(u|x)\, du, \tag{4}$$

assuming that the integral and differential operators are interchangeable. By comparing (3) and (4), it is clear that the effect of x_l on y holding u constant may be very different from the effect of x_l on y not holding u constant. Moreover, in (3) the structural parameters θ have interest *per se*: (i) θ_l gives the direction of the partial effects; (ii) testing the statistical significance of the partial effects is equivalent to test the statistical significance of θ_l; and (iii) the relative effects of changes in any two continuous variables x_l and x_w are given by θ_l/θ_w. However, these three features of θ may be lost when the analysis is conditional only on observables and u and x are not independent. For instance, the partial effect (4) may have a different sign from θ_l; Example 1 (cont.) below provides a very simple illustration of such a case. Hence, in the framework of nonlinear regression models, consistent estimation of parameters and marginal effects are not in general twin issues when the analysis is conditional only on observables.

Example 1 (cont. - Exponential regression model) *Suppose that instead of assuming that $E_u[\exp(u)|x] = 1$, we assume alternatively that u has a normal distribution with mean zero and*

variance ηx_l*. As it is well known, this implies that* $\exp(u)$ *has a log-normal distribution with mean given by* $\exp(0.5\eta x_l)$*. Hence,* $E(y|x) = \exp(x\theta + 0.5\eta x_l)$*,*

$$\frac{\partial y}{\partial x_l} = \theta_l \exp(x\theta + u)$$

and

$$\frac{\partial E(y|x)}{\partial x_l} = (\theta_l + 0.5\eta) \exp(x\theta + 0.5\eta x_l).$$

Thus, both partial effects will have the same sign only if $|\theta_l| > -0.5\eta$*.*

Given the very different implications produced by conditioning or not the analysis on unobservables, it is crucial for empirical researchers to define carefully which are the aims of their analysis in order to decide which approach provides the right answers to their questions. If one is interested in causal effects that respect the notion of a *ceteris paribus* change, then u must be hold constant and, hence, the relevant partial effects are given by (3) and consistent estimation of the structural parameters is essential. In contrast, if the quantities of primary interest for empirical analysis are the partial effects averaged across the distribution of the unobserved heterogeneity, which are given by (4), then the fact that unobserved heterogeneity does not allow identification of the structural parameters in many cases is not necessarily a relevant issue. See Heckman (2001) for a detailed account of both approaches in the context of economic policy evaluation and Wooldridge (2005) for a more general comparison.

A problem with both approaches is that in general it is necessary to make distributional assumptions on the unobservables in order to identify and estimate the structural parameters and/or the partial effects. In Sections 3 and 4 we show that for a particular class of nonlinear regression models such distributional assumptions are not required for either consistent estimation of structural parameters or consistent estimation of (conditional on observables) partial effects.

2.3 GMM estimation

The transformation regression model that is proposed in this paper, see Section 3, is defined by a set of orthogonality conditions between a function of the unobservables, say $u^* \equiv u^*(y, x; \theta)$, and a set of s instrumental variables, $s \geq k$, which we denote by z:

$$E\left(z'u^*\right) = 0. \tag{5}$$

The instruments z may or may not coincide with the explanatory variables, depending on whether the latter variables may be viewed as exogenous or endogenous. As shown later on, u^*

may be a nonlinear function of θ and, therefore, the parameters of interest that appear in (5) have to be estimated by GMM or any other method appropriate for moment condition models. Next, we briefly review the main features of GMM.[3]

Let $m(\cdot) = z'u^*$ be a vector of moment indicators, let $\hat{m}(\theta) \equiv \sum_{j=1}^{N} m(y_j, x_j, z_j; \theta) \Big/ N$, where j indexes each sampling unit and N denotes the number of observations, and let $\hat{W}$ denote a symmetric, positive definite matrix that converges almost surely to a nonrandom, positive definite matrix W. When $s = k$, the typical case where all regressors are exogenous, then, irrespective of the choice of $\hat{W}$, an efficient GMM estimator for θ is given by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{m}(\theta)' \hat{W}^{-1} \hat{m}(\theta), \tag{6}$$

where Θ denotes the parameter space. On the other hand, when $s > k$, a two-step efficient GMM estimator is given by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{m}(\theta)' \hat{\Omega}(\tilde{\theta})^{-1} \hat{m}(\theta), \tag{7}$$

where $\hat{\Omega}(\theta) \equiv \sum_{i=1}^{N} m_i(\theta) m_i(\theta)' \Big/ N$ and $\tilde{\theta}$ is some preliminary estimator defined by an equation similar to (6) with W typically defined as the identity matrix.

Under suitable regularity conditions, see Newey and McFadden (1994), we have in both cases

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left[0, \left(M' \Omega^{-1} M\right)^{-1}\right], \tag{8}$$

where $\xrightarrow{d}$ denotes convergence in distribution, $\Omega \equiv E\left[m(y, x, z; \theta) m(y, x, z; \theta)'\right]$ and $M \equiv E\left[\nabla_{\theta} m(y, x, z; \theta)\right]$.

3 Transformation regression models

The use of transformation regression models is a very common approach in econometrics to deal with unobservables in nonlinear regression models. In this section, we first briefly review the typical transformation that has been applied to most nonlinear regression models. Then, we propose a new transformation regression model that circumvents the inability of the standard approach to accommodate boundary values of the response variable. Both the standard and the new approaches transform the original model (1) in such a way that the coherence with the economic theory that implied equation (1) is kept but orthogonality conditions of the type considered in the previous section, see equation (5), may be straightforwardly generated under

[3] Alternative estimation methods which share the first order asymptotic properties of GMM are those in the generalized empirical likelihood (GEL) class; see *inter alia* Newey and Smith (2004).

the assumption that:

$$E\left(u^*\right| z) = 0. \tag{9}$$

3.1 The standard linearization approach

Assume that there is a monotonic function $H(\cdot) = G(\cdot)^{-1}$ that, applied to both sides of (1), gives rise to the linear model

$$H(y) = x\theta + u. \tag{10}$$

Equations (1) and (10) represent exactly the same deterministic relationship. If u was observed, it would be completely indifferent to work with either equation. However, u is not observed and, hence, given that in (10) the unobservables are additively separable, identification and estimation of θ becomes much simpler when working with model (10). Indeed, assuming that $E(u|z)$ is a constant not depending on z, consistent GMM estimators for the structural parameters are straightforwardly obtained by considering orthogonality conditions generated from (9), with $u^* = u$:

$$E\{[H(y) - x\theta]| z\} = 0. \tag{11}$$

When $z = x$, this corresponds to a simple estimation of (10) by OLS.

The transformation regression model defined by (10) is very simple and may be applied to a wide variety of nonlinear regression models. However, the $H(\cdot)$ function is often not defined for some values of y, as the following examples illustrate.

Example 1 (cont. - Exponential regression model) *In this case, $H(y) = \log(y)$, which gives rise to the well known log-transformed model. Because the log-transformation is not defined for $y = 0$, this approach can only be applied to positive data. Typically, the extensive literature using log-transformed models has overcomed this limitation by adding an arbitrary constant to all observations of y or by dropping observations with $y = 0$. As shown by Santos Silva and Tenreyo (2006), both approaches may originate large biases in the estimation of the parameters of interest.*

Example 2 (cont. - Fractional regression models) *For the models for fractional data described in Table 1, the transformation $H(\cdot)$ is given in Table 2. It is clear that $H(y)$ is not defined for both the boundaries values 0 and 1 in all cases. Therefore, the OLS or GMM estimators based on (10) may only be used in cases where y is observed only on the open interval $]0,1[$. To deal with the boundary values, the same 'solutions' described in the previous example are in general used and the same criticisms apply.*

Table 2 about here

3.2 The proposed transformation

To overcome the problem associated to the linearized model (10), we propose next a different transformation of the specified structural model (1). This new transformation is slightly more complicated than the previous one and gives rise to a regression model that, similarly to the original model, is nonlinear in the parameters. However, the problems with the boundary values that affect the linearized regression model are in general attenuated and, in some cases, even eliminated.

Assume that the function $G(\cdot)$ in (1) may be decomposed as

$$G(x\theta + u) = G_1[G_2(x\theta + u)], \tag{12}$$

where $G_1(\cdot)$ is an invertible function and $G_2(\cdot)$ is a nonlinear function multiplicatively separable into $k + c$ terms, which for our purposes is convenient to group into two terms, one function of $x\theta$ and the other function of u, as follows:

$$G_2(x\theta + u) = G_2(x\theta)\, G_2(u). \tag{13}$$

Typically, $G_2(\cdot)$ will be the exponential function. Assume that $G_2(x\theta) \neq 0$ and that $E[G_2(u)\,|z]$ is a constant not depending on z. In particular, given that x contains a constant term, we may assume without any loss of generality that $E[G_2(u)\,|z] = 1$.[4]

Let $H_1(\cdot) = G_1(\cdot)^{-1}$. Then, model (1) may be written as:

$$\begin{aligned} y &= G_1[G_2(x\theta + u)] \\ H_1(y) &= G_2(x\theta)\, G_2(u). \end{aligned} \tag{14}$$

Dividing both sides of (14) by $G_2(x\theta)$, so that (functions of) x and u become additively separable, and subtracting 1 to both sides of the resultant model produces

$$\frac{H_1(y)}{G_2(x\theta)} - 1 = G_2(u) - 1. \tag{15}$$

As, under the assumptions stated above, $E\{[G_2(u) - 1]\,|z\} = 0$, the left hand-side of this equation may be interpreted as the residual function that appears in (9). Hence, assuming

[4] E.g. if $G_2(\cdot)$ is an exponential function, then we may redefine the constant θ_0 and the error term as $\beta_0 = \exp(\theta_0)\, E[\exp(u)]$ and $\varepsilon = \exp(u)/E[\exp(u)]$), respectively.

standard rank conditions for identification, consistent GMM estimators for θ may be obtained based on a set of orthogonality functions generated from

$$E\left\{\left[\frac{H_1(y)}{G_2(x\theta)}-1\right]\bigg|\, z\right\}=0. \tag{16}$$

The transformation regression model (15) applies to a more restrict class of models than the simple linearized model (10), since it involves more requirements on the definition of $G(\cdot)$. However, as the inverse function $H_1(\cdot)$ is simpler than $H(\cdot)$, see the examples below, less restrictions are created in the domain of y. Indeed, note that while $H_1(\cdot)$ merely transforms $G(\cdot)$ into a (possibly non-linear) function $G_2(\cdot)$ of $x\theta+u$, $H(\cdot)$ goes one step further and reduces $G(\cdot)$ to $x\theta+u$.

While the transformation regression models (10) and (15) represent the same deterministic relationship, their stochastic versions (11) and (16) are not in general equivalent. In particular, the assumptions required for consistent estimation of the parameters of interest in each model do not imply each other. That is, it may be the case that $E(u|z)$ is a constant not depending on z, as required by (11), but that $E[G_2(u)|z]$ is a function of z, unlike required by (16), and vice-versa. Only under the stronger assumption of statistical independence between z and u, which implies that $E[t(u)|z]$ is a constant not depending on z for any function $t(\cdot)$, will the two models produce simultaneously consistent estimators for the slope parameters. Therefore, when application of both transformation regression models is possible, a major discrepancy between the estimated parameters might suggest that some form of misspecification is present in one of the models. The same reasoning applies if a distribution is assumed for u and model (1) is estimated directly.

Example 1 (cont. - Exponential regression model) *In this case, $G_1(\cdot)$ is a linear function, $G_2(\cdot)$ is an exponential function, $H_1(y)=y$ and $\frac{H_1(y)}{G_2(x\theta)}=\frac{y}{\exp(x\theta)}$. This implies that the transformation regression model defined by (15) includes as particular cases two estimators well known in the econometrics literature: the estimator proposed by Mullahy (1997) to deal with endogeneity in count data models; and the Gamma-based QML estimator considered by Manning and Mullahy (2001) and Santos Silva and Tenreyo (2006) for the case of exogenous variables. Clearly, in contrast to the standard linearization approach, in this context there is no problem in dealing with zero outcomes of y.*

Example 2 (cont. - Fractional regression models) *From the four alternative models considered before, transformation (15) may be applied to the logit and complementary loglog models, as described in Table 3, but not to the probit model. Relative to the standard linearization ap-*

proach, this new transformation regression model has the advantage of accommodating one of the two boundary values of fractional responses. This is particularly relevant because many samples cluster only at zero or one (see, for example, the applications by Ramalho and Silva, 2009, and Ramalho, Ramalho and Henriques, 2010, respectively) and not simultaneously at both boundaries. Moreover, note that we can always redefine the response fractional variable and decide to model its complementary, which means that the three models in Table 3 may be used irrespective of the boundary value that is observed in the original sample.

Table 3 about here

3.3 An alternative GMM estimator for the case of endogenous explanatory variables

By defining the composition of the vector z appropriately, the GMM estimator proposed in the previous section is valid under a variety of situations, including cases of endogenous covariates. From now on, we denote by GMM_x the estimator that uses $z = x$ and, thus, is only appropriate when endogeneity is not a problem; and by GMM_z the estimator based on a set of instruments z that were chosen in such a way that the consistency of the parameter estimators is achieved also under endogeneity.

As it is clear from (16), a very attractive feature of the GMM_z estimator is that no assumptions about the reduced form of the endogenous explanatory variables need to be made. Moreover, our estimator applies in exactly the same way irrespective of the endogenous regressors being discrete or continuous. Thus, the GMM_z estimator may be seen as a generalization of Mullahy's (1997) estimator for exponential regression models (see Example 1 (cont.) above) and is in clear contrast to most instrumental variable estimators that have been proposed for nonlinear regression models (e.g. Smith and Blundell, 1986; Rivers and Vuong, 1988; Wooldridge, 1997), which are not robust to misspecification of the reduced form of the endogenous covariates and typically require different procedures according to the characteristics of those variables. Nevertheless, a potentially more efficient estimator may be constructed in the presence of reliable information about the reduced form of the endogenous explanatory variables. Next, we outline how such information may be incorporated in the estimation process of the structural parameters that appear in the transformation regression model (15) in order to obtain an estimator that is similar in spirit to that suggested by Smith and Blundell (1986) for censored regression models, Rivers and Vuong (1988) for binary response models and Wooldridge (1997) for count data / exponential regression models.

Assume that there are k_1 and k_2 exogenous (x_1) and endogenous (x_2) explanatory variables,

respectively. Assume also that strictly monotonic transformations of each x_{2l}, $l = 1, ..., k_2$ can be found so that a linear reduced form with additive disturbances can be found. Let $S(x_2)$ denote the vector of those monotonic transformations. Then, we may write

$$S(x_2) = z\pi + v, \tag{17}$$

where z contains x_1, π is an $s \times k_2$ matrix of reduced form parameters and v is a k_2 vector of reduced form errors. Finally, assume that (u, v) is independent of z and that

$$u = v\rho + \epsilon, \tag{18}$$

where ϵ is independent of v. Under these assumptions, it follows from (1) and (18) that

$$y = G(x\theta + v\rho + \epsilon) \tag{19}$$

and, hence,

$$\frac{H_1(y)}{G_2(x\theta + v\rho)} - 1 = G_2(\epsilon) - 1. \tag{20}$$

Using standard arguments from two-step estimation, it may be shown that GMM estimation based on

$$E\left\{\left[\frac{H_1(y)}{G_2(x\theta + v\rho)} - 1\right] \middle| x, v\right\} = 0, \tag{21}$$

with v replaced by $\hat{v} = S(x_2) - z\hat{\pi}$, where $\hat{\pi}$ is an OLS estimator, produces consistent estimators for θ and ρ; see Newey and McFadden (1994) for the consistency of two-step estimators. Alternatively, we may append the first-order conditions for $\hat{\pi}$ to the moment conditions generated from (21) and estimate simultaneously π, θ and ρ by GMM, which has the advantage of providing directly correct standard-errors to all parameters.

Clearly, unlike the Mullahy-type GMM_z estimator, the consistency of this alternative estimator, denoted from now on by GMM_{xv}, depends crucially on the correctness of both equation (18) and the linear reduced form (17) for (a transformation of) x_2, with z independent of v. However, if both equations are correctly specified, then, by using that extra information, the GMM_{xv} estimator is more efficient. Moreover, testing for endogeneity is simpler in this framework: simply test for the significance of the parameters ρ in (21) using any classical GMM test of parametric restrictions; see *inter alia* Wooldridge (1997) for a discussion of similar tests of endogeneity and Newey and West (1987) for GMM tests of parametric restrictions. In contrast, in the GMM_z framework, it appears that the only form of testing for endogeneity of x_2 would involve the implementation of a standard Hausman test contrasting GMM_x and GMM_z

estimators.

4 Estimation of partial effects conditional on observables

The focus on the previous section was on consistent estimation of the structural parameters. Under the assumptions stated, it is straightforward to obtain consistent estimators of those parameters using the proposed transformation regression model. As estimation is performed in the GMM framework, it is also straightforward to test the significance of structural partial effects and any type of restrictions imposed by economic theory. The estimation of the magnitude of structural partial effects is also very simple: just plug in some values for x and u in (1) or (3). However, while there are many interesting values that can be plugged in for x (e.g. sample means and quartiles), it is less clear which values should be considered for the unobservables. Therefore, many empirical researchers prefer to compute partial effects conditional only on the observables.

As it is clear from (4), estimation of conditional partial effects will require in general making distributional assumptions on the unobservables. Moreover, the integrals that appear in (4), in general, cannot be calculated analytically and have to be computed by numeric integration or simulation. There are, however, some exceptions to this situation. One concerns the exponential regression model, since in this case the structural function $G(\cdot)$ is multiplicatively separable in terms of x and u, see Example 1, and, thus, $f(u|\, x)$ does not need to be specified. The other exceptions concern very special combinations of the structural function and the distribution of unobservables, still requiring making distributional assumptions on unobservables but avoiding the computation of integrals. In fact, when u has a specific distribution (typically, the normal distribution), $G(\cdot)$ has a specific form (e.g. a binary probit model) and u is independent of x, or independent conditionally on a set of additional controls, then estimating the naive model that ignores the presence of unobserved heterogeneity (i.e. estimating the misspecified model $E(y|x) = G(x\theta)$ instead of the correct model defined in (2)), or the model that adds a set of controls to the index function but still omits u, produces biased estimates for the structural parameters but consistent estimators for the partial effects conditional on unobservables; see Wooldridge (2005) for details and examples and Cramer (2007) and Ramalho and Ramalho (2010) for further discussion.

Wooldridge's (2005) approach still requires making distributional assumptions on u, has the undesirable feature (at least for some researchers) of yielding inconsistent estimators for the structural parameters, and is not generally applicable. The transformation regression models proposed in the previous section are also not generally applicable but, in addition to produce

consistent estimators for the structural parameters, have also the ability of generating consistent estimators for conditional partial effects without requiring the full specification of $f(u|\,x)$, provided that we restrict the dependence between observables and unobservables to the conditional mean (i.e. $E\{[G_2(u)-1]\,|x\}$ may depend on x but other functions of u not). Under this additional assumption, the following two-step procedure may be used for estimating partial effects for individual i:

1. Obtain the GMM estimator $\hat{\theta}$ and the residuals $\hat{u}_j^* = \frac{H_1(y_j)}{G_2(x_j\hat{\theta})} - 1$ and $\hat{u}_j = G_2^{-1}\left(\hat{u}_j^* + 1\right)$, $j = 1, ..., N$, where the function G_2 is assumed to be invertible;

2. Compute the partial effect (4) using its sample analog:

$$\left(\widehat{\frac{\partial E(y|x)}{\partial x_l}}\right) = \frac{1}{N}\sum_{j=1}^{N} \frac{\partial\left\{G_1\left[G_2\left(x_i\hat{\theta}\right)G_2(\hat{u}_j)\right]\right\}}{\partial x_l} = \frac{1}{N}\hat{\theta}_l\sum_{j=1}^{N} g\left(x_i\hat{\theta} + \hat{u}_j\right). \qquad (22)$$

Note that this two-step procedure is valid for all the GMM estimators defined in the previous section: GMM_x, GMM_z and GMM_{xv}.

The estimator defined in (22) is a natural extension of the smearing technique suggested by Duan (1983) for the log-transformed model, estimating the unknown error distribution by the empirical distribution function of the GMM residuals calculated in step 1; see also Abrevaya (2002). Although rarely used in the economics literature, this is a very simple method to employ in practice. Of course, the variance of (22) will have to be computed using the delta method or the bootstrap, but that is the standard procedure when working with partial effects in nonlinear models.

5 Monte Carlo simulation study

In this section we carry out a small-scale Monte Carlo study to investigate the finite-sample performance of the estimators proposed in the paper under different simulated scenarios. This study focus on the estimation of the structural parameters of a logit fractional regression model, but considers also the computation of conditional partial effects. For comparative purposes, in addition to the GMM_x, GMM_z and GMM_{xv} estimators, we include also two QML estimators: the standard Bernoulli-based QML estimator proposed by Papke and Wooldridge (1996) for fractional regression models (denoted by QML_x), which does not account for any type of unobserved heterogeneity; and a variant of that estimator (denoted by QML_{xv}), which was proposed by Wooldridge (2005) to deal with endogeneity issues and is constructed in a similar way to the

GMM_{xv} estimator. However, unlike the GMM_{xv} estimator, QML_{xv} does not allow for other sources of heterogeneity.

5.1 Design

All experiments are based on the following structural model:

$$y = G(\theta_0 + \theta_1 x_1 + u), \tag{23}$$

where $G(a) = \exp(a) / [1 + \exp(a)]$, $\theta_0 = 0$, $\theta_1 = 1$ and x_1 denotes a single covariate. The explanatory variable x_1 is generated from either

$$x_1 = z\pi_1 + v \tag{24}$$

or

$$\ln \frac{x_1}{1 - x_1} = z\pi_2 + v, \tag{25}$$

where z is an $(s-1)$-vector of instrumental variables, which are generated as $\mathcal{N}(0,1)$ random variables, with the elements of z independent of each other and of u and v. The reduced form parameters π_1 and π_2 equal an $(s-1)$-vector of ones times a scalar constant Π_1 and Π_2, respectively. We set $s = \{3, 12\}$, $\Pi_1 = 0.5$ and $\Pi_2 = 1$.

We generate the error terms (u, v) as correlated; their joint distribution is $\mathcal{N}(\mu, \Sigma)$, where $\mu = (-0.5, 0)$ and $\Sigma \in \mathcal{R}^{2\times 2}$ with diagonal elements equal to unity and off-diagonal elements ρ_{uv}. Setting the mean of u equal to -0.5 ensures that $E[\exp(u)|z] = 1$, as assumed by our GMM estimators; as pointed out before, a different mean of u would affect only the consistency of the estimator for θ_0. As the error terms have a joint multivariate normal distribution, we may write

$$u = \rho_{uv} v + \epsilon, \tag{26}$$

with ϵ independent of v and z, as in equation (18). Moreover, the variance of ϵ is given by $\sigma_\epsilon^2 = 1 - \rho_{uv}^2$. Hence, while in the $\rho_{uv} = 0$ ($\sigma_\epsilon^2 = 1$) case x_1 is exogenous but there is a large amount of neglected heterogeneity, for $\rho_{uv} = \pm 1$ ($\sigma_\epsilon^2 = 0$) x_1 is strongly endogenous but the impact of neglected heterogeneity is irrelevant (if ignored, only the estimation of the parameter θ_0 is affected). In order to measure the effect of different degrees of endogeneity and neglected heterogeneity over the various estimators we set $\rho_{uv} = \{-1, -0.8, ..., 1\}$.

All experiments were repeated 5000 times using the statistical package R and considering Monte Carlo samples of size $N = \{200, 1000\}$.

5.2 Estimation of structural parameters

Figure 1 presents the results obtained for the first set of experiments, where the reduced form (24) was used both for generating the data and for estimating purposes. For each experiment, we report for the five alternative estimators of the structural parameter θ_1 included in this study the following statistics: the mean across replications; the root mean squared error (RMSE); and the empirical coverage of a 95% confidence interval, which was estimated by taking the proportion of cases where the confidence interval cover the true value of θ_1. In order to estimate correctly the standard errors necessary for the computation of the confidence interval, in the implementation of the QML_{xv} and GMM_{xv} estimators both the parameters of the structural and reduced forms were estimated simultaneously.

The first column of Figure 1 shows clearly that, irrespective of the value of ρ_{uv}, both GMM_{xv} and GMM_z provide consistent estimation of θ_1. In contrast, all the other estimators are biased in most cases. The GMM_x estimator is consistent only for $\rho_{uv} = 0$, its bias increasing as the degree of endogeneity (in absolute value) increases. The QML_{xv} estimator is consistent only when the neglected heterogeneity can be ignored ($\rho_{uv} = \pm 1$), its bias increasing as the variance of v increases (ρ_{uv} decreases), achieving a maximum for $\sigma^2_\epsilon = 1$. That is, the QML_{xv} estimator displays the classic attenuation bias that is often mentioned as the main consequence of neglected heterogeneity. As in all simulated cases there are endogeneity and/or neglected heterogeneity, the standard QML_x estimator, used in most empirical applications, displays large biases most of time, except in a particular situation where the effects of endogeneity and neglected heterogeneity seem to compensate each other. Because they do not account for endogeneity, the bias of both the QML_x and GMM_x estimators may be positive or negative, depending on the value of ρ_{uv}.

Figure 1 about here

The analysis of the RMSE of each estimator shows the importance of using additional information in the estimation process in order to obtain more efficient estimators, especially for smaller sample sizes and when less moment conditions are used.[5] Clearly, in the presence of unobserved heterogeneity, if the empirical researcher knows for sure that endogeneity is not an issue, then he/she should use the GMM_x estimator; if neglected heterogeneity is not a problem and the reduced form of the endogenous explanatory variable is known, then the QML_{xv} estimator is probably the best option; if the data are affected by both neglected heterogeneity and endogeneity issues and the reduced form of the endogenous regressor is known, then it

[5] Note that as π_1 is fixed independently of the number of instruments, more instruments imply a higher overall fit of the instruments to the endogenous regressor x_1.

is preferable to apply the GMM_{xv} estimator. Similar conclusions are achieved if we analyze the graphs relative to the coverage of confidence intervals. The last column of Figure 1 also illustrates the danger of not accounting for the correct type of unobserved heterogeneity: the empirical coverage of the confidence intervals yielded by QML_x, QML_{xv} and GMM_x tend to zero except in the particular cases of neglected heterogeneity (GMM_x) and innocuous neglected heterogeneity (QML_{xv}).

In Figure 2 we consider the case where the reduced form of the endogenous covariate is misspecified. We generate x_1 using the linearized logit model given in (25) but estimate QML_{xv} and GMM_{xv} using the following linearized loglog reduced form:

$$-\ln\left(-\ln x_1\right) = z\pi_2 + v. \tag{27}$$

The results reported in Figure 2 show clearly that, under misspecification of the reduced form, the only estimator that produces reliable estimates of structural parameters is GMM_z. All the other estimators are inconsistent when x_1 is endogenous, with the extent of the bias depending on the degree of endogeneity. In effect, the use of an incorrect reduced form is innocuous for the consistency of the GMM_{xv} estimator only when there is no endogeneity, i.e. precisely in the case where no reduced form for x_1 would need to be specified. In terms of RMSE, the GMM_z estimator displays the most uniform behavior of all estimators, but there are several cases, particularly for $s = 3$ and $N = 200$, where its RMSE is not among the lowest. However, as the sample size grows, the RMSE of the GMM_z estimator decreases substantially, outperforming most of the other estimators also in terms of this criteria when $N = 1000$. Finally, note that, with the exception of the GMM_z estimator, the empirical coverage of a 95% confidence interval converges often to zero for all the other estimators. In particular, note the unreliable behavior of the GMM_{xv} estimator: for some values of ρ_{uv}, that coverage converges to zero as N grows; for other values of ρ_{uv}, it seems to converge to one.

Figure 2 about here

5.3 Conditional partial effects

We now examine the ability of QML and GMM estimators to measure partial effects conditional only on observables. For each GMM estimator, the partial effect is given by $N^{-1}\hat{\theta}_1 \sum_{j=1}^{N} g\left(\hat{\theta}_0 + \hat{\theta}_1\bar{x}_1 + \hat{u}_j\right)$, see equation (22), where $\bar{x}_1$ represents represents one of the $\{0, 0.02, 0.04, ..., 0.98, 1\}$ population quantiles of x_1. To stress the importance of using the proposed two-step procedure for computing conditional partial effects, we estimated also 'naive' partial effects, given simply by

$\hat{\theta}_1 g\left(\hat{\theta}_0 + \hat{\theta}_1 \bar{x}_1\right)$, which sets $u = 0$ in the evaluation of the partial effect. For the QML estimators, we computed also naive partial effects and, only for QML_{xv}, the following smearing-type estimator, suggested by Wooldridge (2005): $N^{-1}\hat{\theta}_1 \sum_{j=1}^{N} g\left(\hat{\theta}_0 + \hat{\theta}_1 \bar{x}_1 + \rho_{uv}\hat{v}_j\right)$. We add the superscript 's' to all estimators that average out the unobservables, e.g. GMM_z^s.

In Figure 3 we display the mean across the replications of estimated partial effects for some selected cases. In particular, only three experimental designs are presented in Figure 3: neglected heterogeneity but no endogeneity ($\rho_{uv} = 0$); strong endogeneity but innocuous neglected heterogeneity ($\rho_{uv} = 1$); and the previous situation but under misspecification of the reduced form. In the three cases, we consider $N = 1000$ and $s = 12$ and display only the partial effects based on the QML_{xv} and GMM_z estimators. As benchmark, we display also the 'true' partial effects, which were calculated by integration as in (4) with $f(u|x)$ replaced by the density used to generate u.

Figure 3 about here

Under neglected heterogeneity and no endogeneity (first graph), note how both the QML_{xv} and QML_{xv}^s estimators yield partial effects very close to the true ones (the same happens with the not reported QML_x estimator), in spite of being based on inconsistent estimates of the structural parameters. This is in accordance with the conjecture by Wooldridge (2005) that when heterogeneity is independent of the covariates and the interest lies in average partial effects of the observed covariates on mean responses, one may simply ignore the unobservables. Wooldridge (2005) demonstrated this result for the probit model with normal-distributed unobservables, but, given the similarity between probit and logit models, it is not surprising that the same conclusion holds approximately for the specification considered in this Monte Carlo study. Regarding the GMM estimators, application of the smearing corrections is clearly essential for estimating consistently conditional partial effects. Otherwise, large biases may be created.

When all relevant heterogeneity concerns the endogeneity of x_1 (second and third graphs), all naive estimators provide biased estimates of partial effects. Applying the smearing correction, the GMM_z^s estimator is the only one that estimates consistently the partial effects in all cases. In the case of QML_{xv}^s (and also GMM_{xv}^s) it is essential to use the right reduced form for x_1.

6 Empirical application: the determinants of corporate capital structure

In this section we use some of the estimators discussed before to assess the determinants of firms' capital structure decisions, namely their option between long-term debt and equity. First, two

competing capital structure theories are briefly discussed, then the main characteristics of the data and variables are described, and finally the main estimation results are presented.

6.1 Capital structure theories

The analysis of the financing decisions of firms has been a key theme in applied corporate finance for more than thirty years. Two of the most popular explanations of firms' debt policy decisions have been the trade-off and the pecking-order theories. According to the former, firms choose the proportion of debt in their capital structure that maximizes their value, balancing the benefits of debt (e.g. the tax deductibility of interest paid) against its costs (e.g. potential bankruptcy costs caused by an excessive amount of debt). In contrast, the pecking-order theory advocates that, due to information asymmetries between firms' managers and potential outside financiers, firms tend to adopt a perfect hierarchical order of financing, giving preference to the use of internal funds and issuing new shares only when their ability to issue safe debt is exhausted. Hence, the firm leverage at each moment merely reflects its external financing requirements without a tendency to revert to any particular capital structure. For details on both theories, see the recent survey by Frank and Goyal (2008).

To evaluate the trade-off and pecking-order theories, many different tests have been proposed in the financial literature. The most common procedure is to use regression models to examine how a given set of potential explanatory variables influences some leverage ratio (e.g. debt to capital or total assets) and then test their significance, examining whether each variable behaves or not as predicted by each theory. Hence, in this framework, the main interest of the econometric analysis lies on the significance of the structural parameters that appear in the leverage equation and not so much on the magnitude of partial effects, be it conditional or not on unobservables. The most common explanatory variables used in capital structure empirical studies, and their effect on leverage ratios as predicted by each capital structure theory, are described in the next section.

6.2 Data and variables

The data set used in this study was provided by the *Banco de Portugal* Central Balance Sheet Data Office and has already been considered by Ramalho and Silva (2009). It comprises financial information and other characteristics of 4692 non-financial Portuguese firms for the year 1999. In accordance with the latest definitions adopted by the European Commission (recommendation 2003/361/EC), each firm is assigned to one of the following two size-based group of firms: micro and small firms; medium and large firms. A separate econometric analysis for each group is

performed.

As a measure of financial leverage, the ratio of long-term debt (defined as the total company's debt due for repayment beyond one year) to long-term capital assets (defined as the sum of long-term debt and equity) is considered (*Leverage*); see Rajan and Zingales (1995) for an extensive discussion on this and other alternative measures of leverage. In all alternative regression models estimated next, the same explanatory variables as those employed by Ramalho and Silva (2009) are contemplated: *Non-debt tax shields* (*NDTS*), measured by the ratio between depreciation and earnings before interest, taxes and depreciation; *Tangibility*, the proportion of tangible assets and inventories in total assets; *Size*, the natural logarithm of sales; *Profitability*, the ratio between earnings before interest and taxes and total assets; *Growth*, the yearly percentage change in total assets; *Age*, the number of years since the foundation of the firm; *Liquidity*, the sum of cash and marketable securities, divided by current assets; and four activity sector dummies: *Manufacturing*; *Construction*; *Trade* (wholesale and retail); and Transport and *Communication*. Some of these variables are expected to have a positive impact on leverage ratios (e.g. *Profitability* and *Liquidity*, in the case of the trade-off theory; *Growth*, in the case of the pecking-order theory; and *Tangibility* and *Size*, in both cases), while others are expected to have a negative effect (e.g. *NDTS* and *Growth*, in the former theory; and *Profitability*, *Age* and *Liquidity*, in the latter); see *inter alia* Ramalho and Silva (2009) for a detailed explanation of these effects.

Table 4 reports descriptive statistics for the dependent and explanatory variables according to the two size-based groups of firms considered in this analysis. Clearly, the group of medium and large firms display a mean leverage ratio that is substantially higher than that of the other group. While this difference may be partially explained by the variables included in the leverage regression, there are many other factors that may affect the capital structure decisions of firms and that, due to data unavailability, typically are not considered in applied work. For example, especially for smaller firms, it is often argued that the personal characteristics of the firms' owners are important factors for explaining firms' financial leverage decisions; see *inter alia* Hutchinson (1995), Vos and Forlong (1996) and Chandler and Hanks (1998). As discussed in previous sections, not accounting for these characteristics may lead to inconsistent estimation of the structural parameters associated to observed explanatory variables and erroneous conclusions about the significance of structural parameters, which is the main interest of the present analysis. As illustrated by the previous example, unobserved heterogeneity may be particularly important for smaller firms, which is the main reason why in this empirical analysis we decided to partition our sample in two groups. Actually, note that even with respect to the observed variables the

smaller firms in our data set are clearly are more heterogenous than larger firms: with the exception of *Age*, all the other explanatory variables display larger standard deviations for the micro and small group of firms.

Table 4 about here

6.3 Econometric analysis

Given that leverage ratios are, by definition, bounded on the closed interval [0,1], several authors have recently explained firms' financing decisions using regression models suitable to deal with fractional responses; see *inter alia* Cook, Kieschnick and McCullough (2008) and Ramalho and Silva (2009). In their formulations, no unobserved heterogeneity is allowed for. Here, we also assume that all observed explanatory variables are exogenous but, in contrast to those authors, allow for the presence of unobservables that are uncorrelated with the observed covariates.

As the minimum value observed for the dependent variable *Leverage* is zero but the maximum is lower than one, see Table 4, any one of the fractional transformation regression models that appear in Table 3 may be used in this context. In contrast, none of the linearized models of Table 2 may be directly applied. Next, we restrict our attention to logit-based regression models, considering the following structural model:

$$y = \frac{\exp\left(x\theta + u\right)}{1 + \exp\left(x\theta + u\right)}. \tag{28}$$

We consider five alternative estimators for (28). The first is the $QMLx$ estimator used by Cook, Kieschnick and McCullough (2008) and Ramalho and Silva (2009), which ignores the presence of unobservables in (28). The second is the $GMMx$ estimator, which yields consistent estimators for the structural parameters under the assumption that $E\left[\exp\left(u\right)|x\right] = 1$. The three other estimators are based on the linearized logit model

$$\ln \frac{y}{1-y} = x\theta + u, \tag{29}$$

dealing with the zero values of y as is typical in log-transformed models. Thus, in one case, all observations with $y = 0$ are dropped, which reduces the sample size by almost 82% (smaller firms) and 50% (larger firms); the resultant estimator is denoted by $OLS\left(y > 0\right)$. In the other cases, we add 0.001 $[OLS\left(y + 0.001\right)]$ or 0.00001 $[OLS\left(y + 0.00001\right)]$ to the value observed for y for all firms. Note that 0.001 is the highest value that we could add to y, given that the maximum value for *Leverage* in the sample is 0.998, see Table 4.

To check the adequacy of each model, we perform a Wald version of the RESET test, which

assesses the correct specification of the conditional expectation assumed in each case by checking the significance of the additional regressor constructed as $\left(x\hat{\theta}\right)^2$. As found out by Ramalho and Ramalho (2012) for binary regression models (which use the same specifications as fractional regression models), the RESET test is sensitive to a large number of model misspecifications, including neglected heterogeneity and heteroskedasticity. All tests were performed using heteroskedasticity-robust estimators.

Table 5 presents the estimation outcomes resulting from the five techniques. The first point to notice is that the truncation applied to y by the $OLS\,(y>0)$ estimator originates in several cases different conclusions from all the other estimators. For instance, the variables *Tangibility* and *Liquidity* for medium and large firms and *Trade* for micro and small firms are important determinants of leverage ratios in all cases except $OLS\,(y>0)$. Conversely, for the latter group of firms, *Growth* and *Age* are relevant covariates only when the model is estimated by $OLS\,(y>0)$. Moreover, note how the effect of the variable *Size* differs dramatically between $OLS\,(y>0)$ and the other estimators: according to $OLS\,(y>0)$, *Size* affects negatively the proportion of debt used by all firms; according to the other estimators, that effect is positive, as predicted by both the trade-off and pecking-order theories. Clearly, the standard approach in many areas of dropping observations not accommodated by the specified model does not seem to be a recommendable practice in the regression analysis of leverage ratios.

Adding a constant to the value observed for y does not seem to be a good idea either. Indeed, although in terms of parameter significance the conclusions produced by both $OLS\,(y+0.001)$ and $OLS\,(y+0.00001)$ are identical, in terms of magnitude there are substantial differences. Typically, the regression coefficients of the latter model are more than 1.5 times the parameter estimates of the former, although in some cases they may be also much lower (e.g. the *Construction* coefficient for micro and small firms). Therefore, as the estimates are very sensitive to the value of the constant added, and this has to be defined in an arbitrary way, application of corrections of this type to overcome the problem of boundary observations is not in general a good option.

The results produced by the $QMLx$ and $GMMx$ estimators are relatively similar in terms of the significance of the parameters. However, particularly for the group of micro and small firms, the same does not happen in terms of the magnitude of the parameters. Moreover, while for the larger group of firms in half of the cases the parameter estimates from $QMLx$ are higher than those from $GMMx$ and in the other half it happens the opposite, for the group of micro and small firms the regression coefficients are systematically much larger (in absolute value) for $GMMx$ (the only exception is the variable *Tangibility*). Given that the most typical

effect of neglected heterogeneity is the production of an attenuation bias in the estimation of regression coefficients, these results suggest that, as anticipated, neglected heterogeneity may be a very important issue in capital structure studies involving small firms and, hence, that the GMM estimators proposed in this paper may be particularly useful in this context. This conjecture is supported also by the RESET test, which in the case of micro and small firms rejects the hypothesis of correct specification of all models except the one that generates the $GMMx$ estimator.

Overall, the results found for the robust $GMMx$ estimator reinforce the conclusion achieved by Ramalho and Silva (2009) that the pecking-order model provides a better explanation of the capital structure decisions of Portuguese firms than the trade-off theory. Indeed, the effects on leverage of the variables *Tangibility* (+), *Size* (+), *Profitability* (-), *Liquidity* (-) and, in the case of larger firms, *Growth* (+) conforms with the former theory and in the last three cases are contrary to the predictions of the latter theory.

7 Conclusion

In this paper we proposed a new set of estimators that are robust to unobserved heterogeneity in a particular class of nonlinear regression models that treat observed and omitted covariates in a similar manner. The suggested estimators are particularly useful when the aim of the analysis is consistent estimation of structural parameters, since they require only a conditional mean assumption regarding a function of the unobservables. For instance, in the framework of fractional regression models, no other method seems to be available to consistently estimate structural parameters under unobserved heterogeneity, unless one is willing to make distributional assumptions on the unobservables and use simulation techniques. Moreover, under some additional assumptions, but still without requiring the full specification of the distribution of the unobservables, our estimators may also be used to estimate partial effects conditional only on observables. One of the estimators proposed has also the very attractive feature of not requiring the specification of a reduced form for the endogenous covariates.

References

Abrevaya, J. (2002), "Computing marginal effects in the Box-Cox model", *Econometric Reviews*, 21(3), 383-393.

Chandler, G.N. and Hanks, S.H. (1998), "An examination of the substitutability of founders human and financial capital in emerging business ventures", *Journal of Business Venturing*,

13, 353-369.

Cook, D.O., Kieschnick, R. and McCullough, B.D. (2008), "Regression analysis of proportions in Finance with self selection", *Journal of Empirical Finance*, 15(5), 860-867.

Cramer, J.S. (2007), "Robustness of logit analysis: unobserved heterogeneity and mis-specified disturbances", *Oxford Bulletin of Economics and Statistics*, 69(4), 545-555.

Duan, N. (1983), "Smearing estimate: a nonparametric retransformation method", *Journal of the American Statistical Association*, 78(383), 605-610.

Frank, M.Z. and Goyal, V.K. (2008), "Tradeoff and pecking order theories of debt", in B.E. Eckbo (ed.) *Handbook of Corporate Finance - Empirical Corporate Finance, Vol. 2*, Elsevier, pp. 135-202.

Heckman, J.J. (2000), "Causal parameters and policy analysis in Economics: a twentieth century retrospective", *Quarterly Journal of Economics*, 115(1), 45-97.

Heckman, J.J. (2001), "Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture", *Journal of Political Economy*, 109, 673-748.

Hutchinson, R.W. (1995), "The capital structure and investment decisions of the small owner-managed firm: some exploratory issues", *Small Business Economics*, 7, 231-239.

Manning, W.G. and Mullahy, J. (2001), "Estimating log models: to transform or not transform?", *Journal of Health Economics*, 20, 461-494.

Mullahy, J. (1997), "Instrumental-variable estimation of count data models: applications to models of cigarette smoking bahavior", *Review of Economics and Statistics*, 79(4), 586-593.

Newey, W. K. and McFadden, D. (1994), "Large sample estimation and hypothesis testing" in R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics, Vol. 4*, Elsevier, pp. 2111-2245.

Newey, W. K., and West, K.D. (1987), "Hypothesis testing with efficient method of moments estimation", *International Economic Review*, 28, 777-787.

Papke, L.E. and Wooldridge, J.M. (1996), "Econometric methods for fractional response variables with an application to 401(k) plan participation rates", *Journal of Applied Econometrics*, 11(6), 619-632.

Rajan, R.J. and Zingales, L. (1995), "What do we know about capital structure? Some evidence from international data", *Journal of Finance*, 50(5), 1421-1460.

Ramalho, E.A. and Ramalho, J.J.S. (2010), "Is neglected heterogeneity really an issue in binary and fractional regression models? A simulation exercise for logit, probit and loglog models", *Computational Statistics and Data Analysis*, 54(4), 987-1001.

Ramalho, E.A. and Ramalho, J.J.S. (2012), "Alternative versions of the RESET test for binary response index models: a comparative study", *Oxford Bulletin of Economics and Statistics*, 74(1), 107-130.

Ramalho, E.A., Ramalho, J.J.S. and Henriques, P. (2010), "Fractional regression models for second stage DEA efficiency analyses", *Journal of Productivity Analysis*, 34(3), 239-255.

Ramalho, J.J.S. and Silva, J.V. (2009), "A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms", *Quantitative Finance*, 9(5), 621-636.

Rivers, D. and Vuong, Q.H. (1988), "Limited information estimators and exogeneity tests for simultaneous probit models", *Journal of Econometrics*, 39(3), 347-366.

Santos Silva, J.M.C. and Tenreyro, S. (2006), "The log of gravity", *Review of Economics and Statistics*, 88(4), 641-658.

Smith, R.J. and Blundell, R.W. (1986), "An exogeneity test for a simultaneous equation tobit model with an application to labor supply", *Econometrica*, 54(3), 679-685.

Vos, E. and Forlong, C. (1996), "The agency advantage of debt over the life cycle of the firm", *Journal of Entrepreneurial & Small Business Finance*, 5(3), 193-211.

Wooldridge, J.M. (1992), "Some alternatives to the Box-Cox regression model", *International Economic Review*, 33(4), 935-955.

Wooldridge, J.M. (1997), "Quasi-likelihood methods for count data", in M.H. Pesaran and P. Schmidt (eds.) *Handbook of Applied Econometrics, Vol. II: Microeconomics*, Blackwell, pp. 352-406.

Wooldridge, J.M. (2005), "Unobserved heterogeneity and estimation of average partial effects", in D.W.K. Andrews and J.H. Stock (eds.) *Identification and Inference for Econometric Models*, Cambridge University Press, pp. 27-55.

Table 1: Alternative fractional regression models

Model	$G(x\theta + u)$
Probit	$\Phi(x\theta + u)$
Logit	$\frac{\exp(x\theta+u)}{1+\exp(x\theta+u)}$
Complementary loglog	$1 - \exp[-\exp(x\theta + u)]$

Table 2: Linearized fractional regression models

Model	$H(y)$
Probit	$\Phi^{-1}(y)$
Logit	$\ln \frac{y}{1-y}$
Complementary log-log	$\ln[-\ln(1-y)]$

Table 3: Transformation fractional regression models

Model	$G_1[G_2(\cdot)]$	$G_2(a)$	$H_1(y)$	$\frac{H_1(y)}{G_2(x\theta)}$
Logit	$\frac{G_2(\cdot)}{1+G_2(\cdot)}$	$\exp(a)$	$\frac{y}{1-y}$	$\frac{y}{1-y}\exp(-x\theta)$
Complementary loglog	$1 - \exp[-G_2(\cdot)]$	$\exp(a)$	$-\ln(1-y)$	$-\ln(1-y)\exp(-x\theta)$

Table 4: Descriptive statistics

Group	Variable	Mean	Min	Max	St.Dev.
Micro and Small	Leverage	0.074	0	0.998	0.189
	NDTS	0.829	0	102.150	3.189
	Tangibility	0.298	0	0.995	0.233
	Size	12.951	7.738	16.920	1.437
	Profitability	0.147	0	1.590	0.117
	Growth	14.923	-81.248	681.354	39.835
	Age	18.267	6	210	12.306
	Liquidity	0.227	0	1	0.247
	Manufacturing	0.563	0	1	0.496
	Construction	0.213	0	1	0.409
	Trade	0.030	0	1	0.171
	Communication	0.116	0	1	0.321
Medium and Large	Leverage	0.148	0	0.978	0.199
	NDTS	0.829	0	26.450	1.479
	Tangibility	0.377	0.002	0.977	0.197
	Size	15.814	11.736	22.270	1.386
	Profitability	0.135	0.001	1.040	0.087
	Growth	8.909	-61.621	188.035	21.014
	Age	28.769	5	184	20.139
	Liquidity	0.120	0	0.963	0.155
	Manufacturing	0.767	0	1	0.423
	Construction	0.121	0	1	0.327
	Trade	0.017	0	1	0.129
	Communication	0.046	0	1	0.210

Table 5: Logit fractional regression models for capital structure decisions

	Micro and small firms					Medium and large firms				
	OLS (y>0)	OLS (y+0.001)	OLS (y+0.00001)	QMLx	GMMx	OLS (y>0)	OLS (y+0.001)	OLS (y+0.00001)	QMLx	GMMx
NDTS	0.0508	-0.0130**	-0.0234**	-0.0316	-0.0961**	-0.0798*	-0.1070***	-0.1726***	-0.1106**	-0.1101
	(0.0368)	(0.0062)	(0.0106)	(0.0314)	(0.0471)	(0.0453)	(0.0352)	(0.0625)	(0.0506)	(0.0769)
Tangibility	0.0241	0.8595***	1.5389***	1.0838***	0.5625	0.3472	3.1608***	5.5869***	1.5296***	1.5626***
	(0.3017)	(0.1984)	(0.3321)	(0.2289)	(0.6183)	(0.2927)	(0.4380)	(0.7531)	(0.2601)	(0.4038)
Size	-0.1035*	0.4523***	0.7818***	0.4757***	0.4861***	-0.1379***	0.3794***	0.7278***	0.1069***	0.1207**
	(0.0535)	(0.0305)	(0.0512)	(0.0379)	(0.1343)	(0.0410)	(0.0600)	(0.1052)	(0.0327)	(0.0570)
Profitability	-3.3667***	-0.9298***	-1.4278***	-2.8171***	-7.3232***	-4.1089***	-5.5156***	-8.8397***	-4.6225***	-6.2095***
	(0.7767)	(0.3108)	(0.5321)	(0.5865)	(1.4259)	(0.6687)	(0.8541)	(1.4932)	(0.6483)	(1.0478)
Growth	0.0039**	-0.0007	-0.0014	-0.0006	-0.0031	0.0073*	0.0133***	0.0212***	0.0069***	0.0121***
	(0.0019)	(0.0009)	(0.0014)	(0.0013)	(0.0030)	(0.0039)	(0.0041)	(0.0067)	(0.0021)	(0.0037)
Age	-0.0145***	0.0018	0.0050	-0.0048	-0.0153	-0.0025	-0.0031	-0.0048	-0.0013	-0.0035
	(0.0050)	(0.0034)	(0.0058)	(0.0042)	(0.0127)	(0.0027)	(0.0039)	(0.0068)	(0.0021)	(0.0030)
Liquidity	-0.1233	-0.5691***	-0.9837***	-1.0443***	-1.8051***	-0.4711	-3.1700***	-5.5614***	-1.8594***	-1.2685**
	(0.3381)	(0.1566)	(0.2640)	(0.3174)	(0.5650)	(0.4273)	(0.4831)	(0.8420)	(0.4023)	(0.6085)
Manufacturing	-0.0427	-0.1374	-0.2240	0.0803	-1.4319*	-0.2382	-0.4247	-0.6476	-0.2709	-0.2322
	(0.2187)	(0.1801)	(0.3004)	(0.1806)	(0.7542)	(0.2091)	(0.3576)	(0.6116)	(0.1917)	(0.2327)
Construction	1.0538***	0.0757	0.0088	0.5332**	-0.7573	0.0918	-0.4349	-0.7943	-0.0598	0.0023
	(0.2708)	(0.2024)	(0.3345)	(0.2103)	(0.8150)	(0.2603)	(0.4444)	(0.7591)	(0.2483)	(0.2987)
Trade	-0.5838	-0.7702***	-1.2566***	-0.7066*	-2.4408***	-0.6743	-1.4412**	-2.5095**	-0.7708	-1.2116**
	(0.4885)	(0.2655)	(0.4508)	(0.3953)	(0.9231)	(0.8856)	(0.6185)	(1.0786)	(0.5415)	(0.5689)
Communication	-0.1954	-0.3083	-0.5255	-0.4514*	-2.5579***	0.1915	-0.5023	-0.9806	-0.0520	0.0484
	(0.2851)	(0.1934)	(0.3261)	(0.2596)	(0.7509)	(0.3155)	(0.5039)	(0.8623)	(0.2635)	(0.3547)
Constant	1.3374	-11.4885***	-19.5024***	-8.7194***	-4.8923***	1.7168**	-9.5426***	-17.2074***	-2.9893***	-2.5685***
	(0.8229)	(0.4349)	(0.7312)	(0.5777)	(1.8807)	(0.7201)	(1.0586)	(1.8557)	(0.5761)	(0.9612)
Number of observations	616	3397	3397	3397	3397	661	1295	1295	1295	1295
RESET p-value	0.0099***	0.0000***	0.0000***	0.0439**	0.2188	0.3637	0.6414	0.7300	0.3267	0.9189

Notes: below the coefficients we report robust standard errors in parentheses; ***, ** and * denote coefficients or test statistics which are significant at 1%, 5% or 10%, respectively.